MassSieve v1.0: A Tool for Parsimony Analysis and Multiple Search Engine Comparisons of LC/MS/MS Proteomics Data

Melinda A. McFarland[1]; Douglas J. Slotta[2]; Sanford P. Markey[1]

[1]National Institute of Mental Health, [2]National Center for Biotechnology Information; NIH, Bethesda, Maryland 20892

Overview

MassSieve has been developed as a platform for parsimony analysis of large scale MS/MS experiments, both for single and multi-comparative analysis • Projects may be saved for future use in MassSieve • Export as CSV

It supports reports from multiple probability-based search engine outputs: Mascot (.dat), Omssa (.omx), and pepXML (X!Tandem, Sequest , Mascot, OMSSA) • Label-free relative quantification information is also available by spectral hit counts per peptide and per protein • Graphical display of each set of related peptides and proteins as well as user defined treatment of indeterminate peptides provides visualization of possible isoforms or conflicting database entries •

Method

Developed in Java and tested on Windows, Mac OSX, and Linux • Uses several open source Java libraries: BioJava to read sequence databases and provide graphical representation of protein sequence coverage, MascotDatfile library to parse Mascot results, Prefuse visualization toolkit to display protein-peptide relationship graphs, GlazedLists framework to display and manipulate tables • LC/MS/MS data searched using Mascot, Sequest, X!Tandem, and OMSSA against the same version of Uniprot SwissProt library • PeptideProphet pepXML results for all algorithms, excluding OMSSA, generated within the LabKey CPAS TPP based pipeline

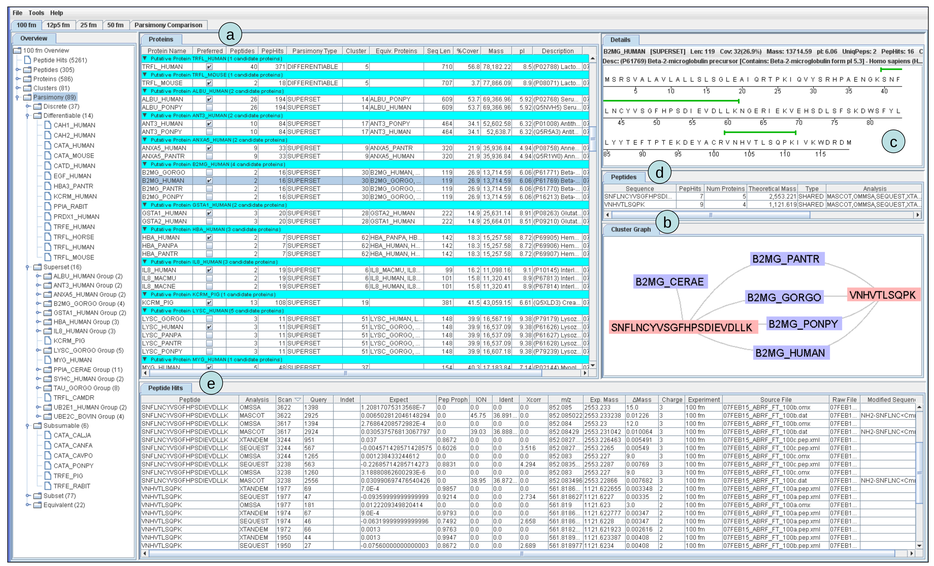


Figure 1. The main window of MassSieve has variable display areas. The tree on the left expands to show peptide and protein lists and parsimony hierarchy. The windows can be moved to display any of the tabs. Selection of a specific peptide or protein triggers related information in corresponding windows.

(a) The parsimony view displays the minimal list of protein associations and corresponding protein centric evidence such as number of unique peptides, number of peptide hits, and sequence coverage.

(b) The cluster graphically displays peptide and protein relations. This permits visual inspection of issues such as isoforms, redundancies, signature peptides, and indeterminate entries.

(c) Loading in the protein sequence fasta file gives access to protein sequence in the details window. Green bars denote identified peptides.

(d) The peptide view is a non-redundant list of all peptides. Can be displayed on a per protein or total experiment basis. A total protein list can also be displayed.

(e) Peptide hits can be displayed specific to a protein, or total experiment.

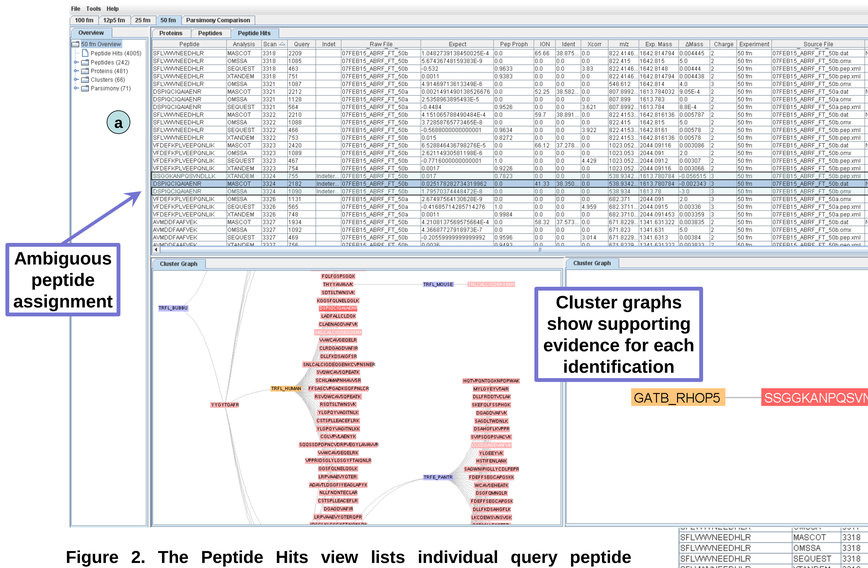
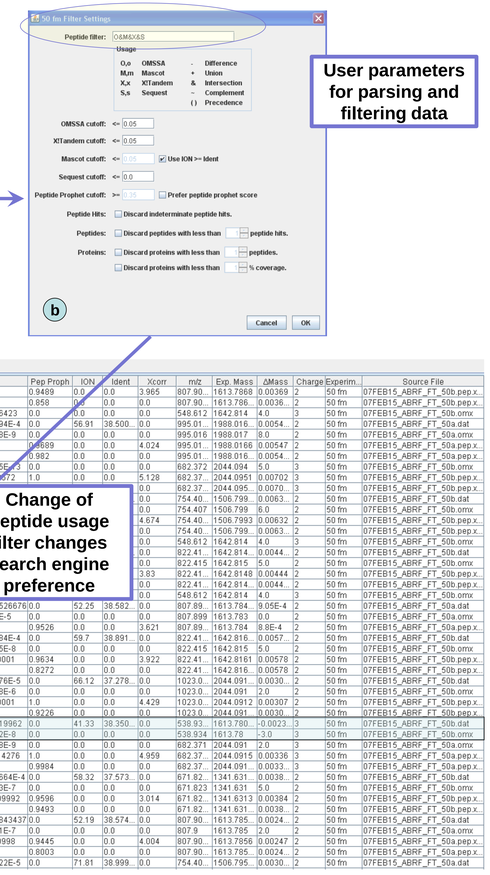


Figure 2. The Peptide Hits view lists individual query peptide assignments.

(a) Sorting by scan number provides a view of ambiguous or missing assignments. This view is particularly useful for evaluating changes in data analysis parameters and search algorithms. Search score filters can be adjusted on the fly to tune score thresholds across scoring platforms.

(b) Peptide filter can be changed for an assortment of multi-search engine uses. Here we include only the intersection of peptide identifications. Note the 'one hit wonder' ambiguous assignment with no supporting evidence has been filtered out.



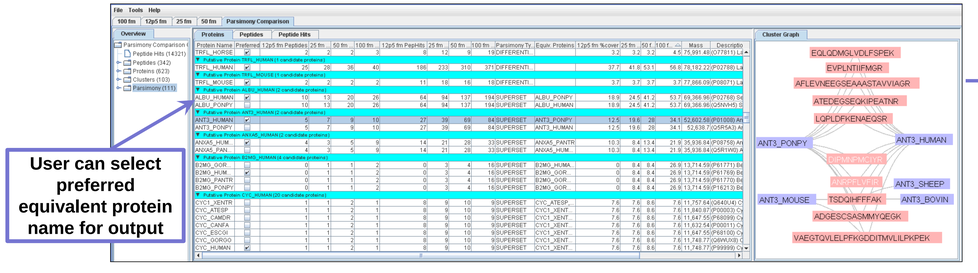


Figure 3. The parsimony comparison report is the most useful for analyzing differences between experiments. Here we assume that experiments are related, such that evidence across all experiments can be used to support parsimony analysis. While parsimony is for the merged data, individual experiments remain distinct for the reporting of numbers of peptides and peptide hits. This is directly applicable to peptide hit based relative quantification.

Figure 4. All reports can be output in several combinations as CSV. For publication we provide an output that contains parsimonious proteins, their associated peptides, numbers of hits, and distribution across each experiment.

To Do

Integration of ion-current extraction utility • Links to MSMS spectral assignments for manual validation • Individual query rejection for parsimony analysis post validation • Integration of protein and gene based associations, such as GO terms and gene symbols • Integrate with the CPAS LIMS • Comparison of parsimony versus protein level scoring

Open source software available at: www.proteomecommons.org/dev/masssieve/

This research was supported in part by the Intramural Program of National Institutes Health, NIMH Z01 MH000279.